



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification<sup>4</sup> :</b>  <b>C12N 15/00, 9/08</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 89/ 03424</b>  <b>(43) International Publication Date:</b> 20 April 1989 (20.04.89)
<b>(21) International Application Number:</b> PCT/GB88/00833 <b>(22) International Filing Date:</b> 7 October 1988 (07.10.88)  <b>(31) Priority Application Number:</b> 8723662 <b>(32) Priority Date:</b> 8 October 1987 (08.10.87) <b>(33) Priority Country:</b> GB  <b>(71) Applicant (for all designated States except US):</b> BRITISH BIO-TECHNOLOGY LIMITED [GB/GB]; Brook House, Watlington Road, Cowley, Oxford OX4 5LY (GB).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> EDWARDS, Richard, Mark [GB/GB]; 7 Ludlow Drive, Thame, Oxon (GB). BURKE, Julian, Francis [GB/GB]; 6 Friar's Walk, Lewes, East Sussex BN7 2LE (GB).		<b>(74) Agents:</b> SHEARD, Andrew, Gregory et al.; Kilburn & Strode, 30 John Street, London WC1N 2DD (GB).  <b>(81) Designated States:</b> AT (European patent), BE (European patent), CH (European patent), DE (European patent), DK, FI, FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), NO, SE (European patent), US.  <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> SYNTHETIC GENE  <b>(57) Abstract</b>  Synthetic DNA coding for horseradish peroxidase includes the sequence (I), and incorporates useful restriction sites at frequent intervals to facilitate the cassette mutagenesis of selected regions. Also included are flanking restriction sites to simplify the incorporation of the gene into any desired expression system.		

ATG CAG TTA ACC CCT ACA TTC TAC GAC AAT AGC TGT  
 CCC AAC GTG TCC AAC ATC GTT CGC GAC ACA ATC GTC  
 AAC GAG CTC AGA TCC GAT CCC AGG ATC GCT GCT TCA  
 ATA TTA CGT CTG CAC TTC CAT GAC TGC TTC GTG AAT  
 GGT TGC GAC GCT AGC ATA TTA CTG GAC AAC ACC ACC  
 AGT TTC CGC ACT GAA AAG GAT GCA TTC GGG AAC GCT  
 AAC AGC GCC AGG GGC TTT CCA GTG ATC GAT CGC ATG  
 AAG GCT GCC GTT GAG TCA GCA TGC CCA CGA ACA GTC  
 AGT TGT GCA GAC CTG CTG ACT ATA GCT GCG CAA CAG  
 AGC GTG ACT CTT GCA GGC GGA CCG TCC TGG AGA GTG  
 CCG CTC GGT CGA CGT GAC TCC CTA CAG GCA TTC CTA  
 GAT CTG GCC AAC GCC AAC TTG CCT GCT CCA TTC TTC  
 ACC CTG CCC CAG CTG AAG GAT AGC TTT AGA AAC GTG  
 GGT CTG AAT CGC TCG AGT GAC CTT GTG GCT CTG TCC  
 GGA GGA CAC ACA TTT GGA AAG AAC CAG TGT AGG TTC  
 ATC ATG GAT AGG CTC TAC AAT TTC AGC AAC ACT GGG  
 TTA CCT GAC CCC ACG CTG AAC ACT ACG TAT CTC CAG  
 ACA CTG AGA GGC TTG TGC CCA CTG AAT GGC AAC CTC  
 AGT GCA CTA GTG GAC TTT GAT CTG CGG ACC CCA ACC  
 ATC TTC GAT AAC AAG TAC TAT GTG AAT CTA GAG GAG  
 CAG AAA GGC CTG ATA CAG AGT GAT CAA GAA CTG TTT  
 AGC AGT CCA AAC GCC ACT GAC ACC ATC CCA CTG GTG  
 AGA AGT TTT GCT AAC TCT ACT CAA ACC TTC TTT AAC  
 GCC TTC GTG GAA GCC ATG GAC CGT ATG GGT AAC ATT  
 ACC CCT CTG ACG GGT ACC CAA GGC CAG ATT CGT CTG  
 AAC TGC AGA GTG GTC AAC AGC AAC TCT TAA

(I)

***FOR THE PURPOSES OF INFORMATION ONLY***

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	ML	Mali
AU	Australia	GA	Gabon	MR	Mauritania
BB	Barbados	GB	United Kingdom	MW	Malawi
BE	Belgium	HU	Hungary	NL	Netherlands
BG	Bulgaria	IT	Italy	NO	Norway
BJ	Benin	JP	Japan	RO	Romania
BR	Brazil	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	LI	Liechtenstein	SN	Senegal
CH	Switzerland	LK	Sri Lanka	SU	Soviet Union
CM	Cameroon	LU	Luxembourg	TD	Chad
DE	Germany, Federal Republic of	MC	Monaco	TG	Togo
DK	Denmark	MG	Madagascar	US	United States of America
FI	Finland				

SYNTHETIC GENE

This invention relates to synthetic genes coding for horseradish peroxidase.

Horseradish peroxidase C (E.C.1.11.1.7) (HRP) is the major peroxidase isozyme isolated from the horseradish (Armoracia rusticana). It is a monomeric glycoprotein of 308 amino acids the polypeptide chain having a MW of 33,980 D. There are three neutral carbohydrate side chains and 4 disulphide bridges. The amino acid sequence of the mature protein has been determined. The presence of a pyrrolidonecarboxyl amino terminus indicates that the protein is probably produced as a precursor form that is processed on secretion. The active form of the enzyme contains a hemin prosthetic group.

The enzyme is particularly stable and is amenable to crosslinking and derivitisation without excessive loss of activity. This together with its wide range of chromogenic substrates, some of which give rise to insoluble, chemiluminescent or fluorescent products, and the low background activities observed in most applications, have made horseradish peroxidase an invaluable tool for diagnostic and research applications in the fields of immunology, histochemistry, cytology and molecular biology. A further advantage it presents over other enzymatic markers is that some substrates for the enzyme give rise to electron dense products that allow correlation of peroxidase location with cellular

ultrastructure using electron microscopy. In addition, horseradish peroxidase is electron dense itself by virtue of the Fe it contains and as a result can act as an E.M. marker in its own right. Particular applications have been found in immunochemistry, where peroxidase cross linked to immunoglobulin is widely used in both ELISA based assay systems and immunocytochemistry. Methods have been described that use either direct crosslinking of peroxidase to the immunoglobulin or indirect crosslinking of biotin labelled immunoglobulin to a streptavidin/horseradish peroxidase complex. Such streptavidin complexes have also found widespread application in nucleic acid hybridisation methods where biotinylated probe sequences can be localised by sequential incubation with the streptavidin/peroxidase complex and a suitable chromogenic peroxidase substrate.

The amino acid sequence of horseradish peroxidase is taught by Welinder, K.G. (Eur. J. Biochem. 96, 483-502 (1979)). The cloning of the cDNA or natural gene for horseradish peroxidase has not been described.

In order to facilitate the dissection of the structure/function relationships of HRP, its incorporation into expression vectors and the production of novel chimeric proteins containing HRP functionality an improved novel synthetic gene for the peroxidase C produced by Armoracia rusticana is sought.

It is by no means easy to predict the design of an improved HRP gene, since the factors that determine the

expressibility of a given DNA sequence are still poorly understood. Furthermore, the utility of the gene in various applications will be influenced by such considerations as codon usage and restriction sites. The present invention relates to a synthetic HRP gene which has advantages in the ease with which it can be modified due to the presence of useful restriction sites.

When synthesising and assembling genes, problems have been encountered when there are inverted or direct repeats greater than eight bases long in the genetic sequence. In addition, areas of unbalanced base composition such as G/C or A/T rich regions or polypurine/polypyrimidine tracts have been found to lead to inefficient expression. The present invention seeks to overcome or at least alleviate these difficulties.

According to a first aspect of the invention, there is provided DNA coding for HRP and having restriction sites for the following enzymes:

HpaI, SacI, SspI, NheI, NsiI, ClaI, PvuI, SphI,  
BspMI, FspI, RsrII, SalI, BglII, BalI, PvuII,  
XhoI, BspMII, BstEII, SnaBI, PflMI, ApaLI, SpeI,  
ScaI, XbaI, StuI, BclI, BstXI, NcoI, KpnI, and  
PstI.

The DNA may also contain a 5' HindIII site and/or a 5' NdeI site and/or a 3' BamHI site and/or a 3' EcoRI site.

According to a second aspect of the invention, there is provided DNA including the following sequence:

CAG TTA ACC CCT ACA TTC TAC GAC AAT AGC TGT CCC  
AAC GTG TCC AAC ATC GTT CGC GAC ACA ATC GTC AAC  
GAG CTC AGA TCC GAT CCC AGG ATC GCT GCT TCA ATA  
TTA CGT CTG CAC TTC CAT GAC TGC TTC GTG AAT GGT  
TGC GAC GCT AGC ATA TTA CTG GAC AAC ACC ACC AGT  
TTC CGC ACT GAA AAG GAT GCA TTC GGG AAC GCT AAC  
AGC GCC AGG GGC TTT CCA GTG ATC GAT CGC ATG AAG  
GCT GCC GTT GAG TCA GCA TGC CCA CGA ACA GTC AGT  
TGT GCA GAC CTG CTG ACT ATA GCT GCG CAA CAG AGC  
GTG ACT CTT GCA GGC GGA CCG TCC TGG AGA GTG CCG  
CTC GGT CGA CGT GAC TCC CTA CAG GCA TTC CTA GAT  
CTG GCC AAC GCC AAC TTG CCT GCT CCA TTC TTC ACC  
CTG CCC CAG CTG AAG GAT AGC TTT AGA AAC GTG GGT  
CTG AAT CGC TCG AGT GAC CTT GTG GCT CTG TCC GGA  
GGA CAC ACA TTT GGA AAG AAC CAG TGT AGG TTC ATC  
ATG GAT AGG CTC TAC AAT TTC AGC AAC ACT GGG TTA  
CCT GAC CCC ACG CTG AAC ACT ACG TAT CTC CAG ACA  
CTG AGA GGC TTG TGC CCA CTG AAT GGC AAC CTC AGT  
GCA CTA GTG GAC TTT GAT CTG CGG ACC CCA ACC ATC  
TTC GAT AAC AAG TAC TAT GTG AAT CTA GAG GAG CAG  
AAA GGC CTG ATA CAG AGT GAT CAA GAA CTG TTT AGC  
AGT CCA AAC GCC ACT GAC ACC ATC CCA CTG GTG AGA  
AGT TTT GCT AAC TCT ACT CAA ACC TTC TTT AAC GCC  
TTC GTG GAA GCC ATG GAC CGT ATG GGT AAC ATT ACC  
CCT CTG ACG GGT ACC CAA GGC CAG ATT CGT CTG AAC  
TGC AGA GTG GTC AAC AGC AAC TCT

The above sequence may be immediately preceeded by an initiation codon (ATG) and immediately followed by a

termination codon (TAA), but this will not necessarily be the case if the DNA incorporates linker(s) and/or extension(s), such as a sequence coding for a signal peptide, for example for efficient expression in eukaryotic cells such as mammalian cells. One extension which gives good expression in mammalian cells is a 5'-extension coding for the amino acids KCSWVIFFLMAVVTGVNS, which may be provided between an initiation codon and the codon coding for the first Q residue. A preferred such extension is shown in Figure 6. A sequence coding for a 3'-signal sequence may code for LLHDMVEVVDFVSSM; a preferred DNA sequence coding for this series of amino acid residues is also shown in Figure 6.

A synthetic HRP gene as described above incorporates useful restriction sites at frequent intervals to facilitate the cassette mutagenesis of selected regions. Also included in preferred embodiments are flanking restriction sites to simplify the incorporation of the gene into any desired expression system.

Codons are those that are favoured by E. coli but it is expected that the DNA would be suitable for expression in other organisms including yeast and mammalian cells.

According to a third aspect of the invention, there is provided a genetic construct comprising DNA according to the first or second aspect or a fragment thereof. The fragment may comprise at least 10, 20, 30, 40 or 50 nucleotides. A genetic construct in accordance with the

third aspect may be a vector, such as a plasmid, cosmid or phage.

According to a fourth aspect of the invention, there is provided a process for the preparation of DNA in accordance with the first or second aspect or a genetic construct in accordance with the third aspect, the process comprising coupling successive nucleotides and/or ligating appropriate oligomers.

The invention also relates to other nucleic acid (including RNA) either corresponding to or complementary to DNA in accordance with the first or second aspects.

The invention encompasses a process for the production of monodisperse horseradish peroxidase C comprising the expression of at least part of a genetic construct as described above.

Further, the invention extends to constructs as described above comprising all or a fragment of a sequence in accordance with the first or second aspect fused to any other sequence of DNA so as to result in a sequence capable of encoding a hybrid protein possessing peroxidase activity. An example of such a construct is a genetic fusion between a gene encoding horseradish peroxidase and a gene encoding streptavidin or avidin such that the encoded fusion protein possesses both biotin binding and peroxidase activity. Another example is a genetic fusion between a gene encoding horseradish peroxidase and a gene encoding an



immunoglobulin-derived antigen binding function such that the fusion protein possesses both antigen binding and horseradish peroxidase activity. The antigen binding function may be an immunoglobulin heavy chain or light chain or fragments thereof or an engineered monomeric antigenic recognition site.

Particular constructs of interest include: vectors comprising the gene for horseradish peroxidase C that enable the production of fusions between horseradish peroxidase and any other protein of interest; and expression vectors that provide for the co-expression of the gene for horseradish peroxidase and another gene of interest either as a single fusion product, as a single polycistronic message or as two separate but linked transcriptional units.

According to a further aspect of the invention, there is provided a gene for horseradish peroxidase containing a mutation (either missense, nonsense, deletion, insertion, duplication or other rearrangement) that destroys or impairs the activity of the encoded horseradish peroxidase protein. The invention extends to genetic constructs including all or a fragment of such a mutant horseradish peroxidase gene.

Defective or non-defective horseradish peroxidase genetic constructs can be employed (for example as markers) in mammalian cells and/or in transgenic animals.

Specific applications of synthetic genes for horseradish peroxidase, which themselves form further aspects of the invention, are disclosed in greater detail below:

1) The gene can be incorporated into a suitable expression vector to allow for the efficient production of the enzyme in a compatible organism. This will have the advantage of being a ready source of a monodisperse enzyme preparation free of the contaminating isozymes present in the material isolated from horseradish root. Varying the organism or cell type chosen for production will also allow for the production of HRP with different patterns of glycosylation, including no glycosylation. Such material will have better defined properties that will make it more suitable for more demanding histochemical applications and sensitive enzyme assays, especially immunoassays.

2) The gene can be incorporated into an HRP-streptavidin or HRP-avidin gene fusion. This will allow for the production of streptavidin-HRP or avidin-HRP complexes without the need for cross-linking. Again this will allow for a better defined, more stable product and will probably result in less loss of both biotin binding and peroxidase activity.

3) Similarly, fusions between immunoglobulins and HRP or protein A and HRP can be produced that would be valuable histochemical reagents. Again the need for the usual cross-linking procedures would be avoided.

4) The HRP gene would have valuable applications in the construction of vectors designed to allow the production of fusions between HRP and any other protein for which a gene or cDNA had been cloned or for which the amino acid sequence is known. This would be useful both for monitoring the expression of a gene the product of which is difficult to assay and to tag the protein of interest to allow its metabolism and pharmacodynamics to be followed in vivo by the use of the appropriate histochemical techniques or enzyme assays. Additionally, HRP fusions will allow for a simple immunopurification of the fusion product through the use of an appropriate anti-HRP antibody.

5) The expression of HRP will be a useful marker in expression systems, eg mammalian cell expression systems. The HRP gene could be expressed either as a fusion or on a polycistronic message with the gene of interest, or as a separate but closely linked transcriptional unit. The production of the easily assayed HRP could be readily screened for and used as an indication as to which clones of cells were likely to be expressing large quantities of the desired product. The use of fluorescent or chemiluminescent HRP chromogenic substrates would allow for the possibility of directly selecting high producing eukaryotic cells by fluorescence activated cell-sorting (FACS).

6) HRP genes carrying mutations (missense, nonsense, deletion, insertion, duplication or other rearrangement) that destroy or impair the enzymatic

activity of the resultant product would allow the construction of vectors that could be used to follow the frequency of reversion or suppression of the particular mutation introduced into the gene.

The introduction of such defective HRP genes into the germ line of the organism of interest would also enable a researcher to fate-map particular cell-types by histologically examining the pattern of HRP activity in the tissue of interest. Care would have to be exercised in constructing a mutant HRP gene with the correct in vivo reversion rate so that areas of HRP activity and hence the presence of reverted HRP gene could be taken as evidence for the clonal origin of the HRP+ cells. The intact synthetic non-mutant gene could also be used for such fate-mapping experiments by infection of an organism with the HRP gene in a suitable vector such as a retroviral vector or transposon.

7) The advantage of a synthetic gene for HRP allows for the production of HRP genes modified to encode a protein carrying small additional sequences, such as N- or C- terminal extensions. These will be of great application in simplifying the purification of the HRP and/or increasing the ease and enhancing the specificity with which it can be cross-linked to other proteins of interest or otherwise derivatised. For example, a C-terminal extension of six to eight Arg residues could be used to simplify purification by analogy with the technique of Sassenfeld *et al.* Bio/technology 2 76 (1984). Alternatively, a tail of

Lys residues would provide an accessible and sensitive site for reaction with bifunctional cross-linking reagents such as glutaraldehyde.

Preferred embodiments and examples of the invention will now be described. In the following description, reference is made to a number of drawings, in which:

- Figure 1            shows the amino acid sequence of horseradish peroxidase C;
- Figure 2            shows the sequence of the horseradish peroxidase synthetic gene; a summary of useful restriction sites; and a sequence of front and back halves of the gene that were initially cloned;
- Figure 3            shows a sequence of synthetic horseradish peroxidase gene divided into oligonucleotides;
- Figure 4            shows a summary of assembly procedure used;
- Figure 5            shows the structure of the HRP E. coli expression plasmid pSD18;
- Figure 6            shows a synthetic HRP gene modified for efficient expression in mammalian cells; and
- Figure 7            shows the structure of the HRP mammalian expression plasmid pCP21.

Example 1

The gene was designed to be synthesised and cloned, in this example, in two halves with a final sub-cloning step to yield the full length gene. The sequence of the two halves of the gene together with that of the final product are depicted in Figure 2. The final synthetic gene encodes the entire mature horseradish peroxidase protein together with the required initiator methionine residue but lacks the leader sequence that is assumed to be present in the natural gene. It is envisaged that the leader sequence appropriate to the expression system of choice would be added to the synthetic gene as required or omitted to allow for intracellular expression of the gene.

The desired gene sequence was divided into a front half and a back half of 501 and 474 bp respectively. Both halves were designed with a common XhoI site to allow for the complete gene to be assembled with a simple cloning step. The front and back halves of the gene were divided into 24 and 22 oligodeoxyribonucleotides (oligomers) respectively as depicted in Figure 3. The division was such as to provide 7 base cohesive ends after annealing complementary pairs of oligomers. The end points of the oligomers were chosen to minimise the potential for inappropriate ligation of oligomers at the assembly stage.

The oligomers were synthesised by automated solid phase phosphoramidite chemistry. Following de-blocking and removal from the controlled pore glass support the oligomers were purified on denaturing polyacrylamide

gels, further purified by ethanol precipitation and finally dissolved in water prior to estimation of their concentration.

All the oligomers with the exception of the 5' terminal oligomers BB279 and BB302 for the front half and BB303 and BB324 for the back half were then kinased to provide them with a 5' phosphate as required for the ligation step. Complementary oligomers were then annealed and the oligomers ligated together by T4 DNA ligase as depicted in Figure 4. The ligation products were separated on a 2% low gelling temperature (LGT) gel and the bands corresponding to the front and back halves of the horseradish peroxidase gene were cut out and extracted from the gel. The purified fragments were then ligated separately to EcoRI/HindIII cut DNA of the plasmid vector pUC18. The ligated products were transformed into HW87 and plated on L-agar plates containing 100 mcg ml<sup>-1</sup> ampicillin. Colonies containing potential clones were then grown up in L-broth containing ampicillin at 100 mcg ml<sup>-1</sup> and plasmid DNA isolated. Positive clones were identified by direct dideoxy sequence analysis of the plasmid DNA using the 17 base universal primer, a reverse sequencing primer complementary to the opposite strand on the other side of the polylinker and some of the oligomers employed in the assembly of the gene that served as internal primers. One front half and one back half clone were subsequently re-sequenced on both strands to confirm that no mutations were present. The complete gene was then assembled by isolating the 466 bp XhoI-EcoRI fragment from the back half clone that contained the 3' end of the gene and ligating it to a

front half clone that had also been digested with EcoRI and XhoI. The identity of the final construct was confirmed by restriction analysis and subsequent complete resequencing.

All the techniques of genetic manipulation used in the manufacture of this gene are well known to those skilled in the art of genetic engineering. A description of most of the techniques can be found in one of the following laboratory manuals: Molecular Cloning by T. Maniatis, E.F. Fritsch and J. Sambrook published by Cold Spring Harbor Laboratory, Box 100, New York, USA, or Basic Methods in Molecular Biology by L.G. Davis, M.D. Dibner and J.F. Battey published by Elsevier Science publishing Co. Inc. New York, USA.

Additional and modified methodologies are detailed below.

#### 1) Oligonucleotide synthesis

The oligonucleotides were synthesised by automated phosphoramidite chemistry using cyanoethyl phosphoramidites. The methodology is now widely used and has been described (Beaucage, S.L. and Caruthers, M.H. Tetrahedron Letters. 24, 245 (1981)).

#### 2) Purification of Oligonucleotides

The oligonucleotides were de-protected and removed from the CPG support by incubation in concentrated NH<sub>3</sub>. Typically, 50 mg of CPG carrying 1 micromole of oligonucleotide was de-protected by incubation for 5 hr



at 70° in 600 mcl of concentrated NH<sub>3</sub>. The supernatant was transferred to a fresh tube and the oligomer precipitated with 3 volumes of ethanol. Following centrifugation the pellet was dried and resuspended in 1 ml of water. The concentration of crude oligomer was then determined by measuring the absorbance at 260 nm.

For gel purification 10 absorbance units of the crude oligonucleotide were dried down and resuspended in 15 mcl of marker dye (90% de-ionised formamide, 10mM tris, 10 mM borate, 1mM EDTA, 0.1% bromophenol blue). The samples were heated at 90° for 1 minute and then loaded onto a 1.2 mm thick denaturing polyacrylamide gel with 1.6 mm wide slots. The gel was prepared from a stock of 15% acrylamide, 0.6% bisacrylamide and 7M urea in 1 X TBE and was polymerised with 0.1% ammonium persulphate and 0.025% TEMED. The gel was pre-run for 1 hr. The samples were run at 1500 V for 4-5 hr. The bands were visualised by UV shadowing and those corresponding to the full length product cut out and transferred to micro-testubes. The oligomers were eluted from the gel slice by soaking in AGEB (0.5 M ammonium acetate, 0.01 M magnesium acetate and 0.1 % SDS) overnight. The AGEB buffer was then transferred to fresh tubes and the oligomer precipitated with three volumes of ethanol at -70° for 15 min. The precipitate was collected by centrifugation in an Eppendorf microfuge for 10 min, the pellet washed in 80 % ethanol, the purified oligomer dried, redissolved in 1 ml of water and finally filtered through a 0.45 micron micro-filter. The concentration of purified product was measured by determining its absorbance at 260 nm.

### 3) Kinasing of oligomers

250 pmole of oligomer was dried down and resuspended in 20 mcl kinase buffer (70 mM Tris pH 7.6, 10 mM MgCl<sub>2</sub>, 1 mM ATP, 0.2 mM spermidine, 0.5 mM dithiothreitol). 10 u of T4 polynucleotide kinase was added and the mixture incubated at 37° for 30 min. The kinase was then inactivated by heating at 85° for 15 min.

### 4) Annealing

8 mcl of each oligomer was mixed, heated to 90° and then slow cooled to room temperature over a period of an hour.

### 5) Ligation

5 mcl of each annealed pair of oligomers were mixed and 10 X ligase buffer added to give a final ligase reaction mixture (50 mM Tris pH 7.5, 10 mM MgCl<sub>2</sub>, 20 mM dithiothreitol, 1 mM ATP. T4 DNA ligase was added at a rate of 100 u per 50 mcl reaction and ligation carried out at 15° for 4 hr.

### 6) Agarose gel electrophoresis

Ligation products were separated using 2% low gelling temperature agarose gels in 1 X TBE buffer (0.094 M Tris pH8.3, 0.089 M boric acid, 0.25 mM EDTA) containing 0.5 mcg ml<sup>-1</sup> ethidium bromide.

### 7) Isolation of ligation products

The band corresponding to the expected horseradish peroxidase gene or gene fragment ligation product was identified by reference to size markers under long wave UV illumination. The band was cut out of the gel and the DNA extracted as follows.

The volume of the gel slice was estimated from its weight and then melted by incubation at 65° for 10 min. The volume of the slice was then made up to 400 µl with TE (10 mM Tris pH 8.0, 1 mM EDTA) and Na acetate added to a final concentration of 0.3 M. 10 µg of yeast tRNA was also added as a carrier. The DNA was then subjected to three rounds of extraction with equal volumes of TE equilibrated phenol followed by three extractions with ether that had been saturated with water. The DNA was precipitated with 2 volumes of ethanol, centrifuged for 10 min in a microfuge, the pellet washed in 70 % ethanol and finally dried down. The DNA was taken up in 20 µl of TE and 2 µl run on a 2 % agarose gel to estimate the recovery of DNA.

### 8) Cloning of fragments

For the initial cloning of the two halves of horseradish peroxidase 0.5 µg of pUC18 DNA was prepared by cleavage with HindIII and EcoRI as advised by the suppliers. The digested DNA was run on an 0.8 % LGT gel and the vector band purified as described above. For the final assembly step the clone carrying the front half of the horseradish peroxidase gene was treated similarly using the enzymes XhoI and EcoRI.

20 ng of cut vector DNA was then ligated to various peroxidase gene DNA ranging from 2 to 20 ng for 4 hr using the ligation buffer described above. The ligation products were used to transform competent HW87 as has been described. Ampicillin resistant transformants were selected on L-agar plates containing 100 mcg ml<sup>-1</sup> ampicillin.

#### 9) Isolation of plasmid DNA

Plasmid DNA was prepared from the colonies containing potential horseradish peroxidase clones essentially as described (Ish-Horowicz, D., Burke, J.F. Nucleic Acids Research 9 2989-2998 (1981)).

#### 10) Dideoxy sequencing

The protocol used was essentially as has been described (Biggin, M.D., Gibson, T.J., Hong, G.F. P.N.A.S. 80 3963-3965 (1983)). The method was modified to allow sequencing on plasmid DNA as described (Guo, L-H., Wu, R. Nucleic Acids Research 11 5521-5540 (1983)).

#### 11) Transformation

Transformation was accomplished using standard procedures. The strain used as a recipient in the cloning was HW87 which has the following genotype:

araD139(ara-leu)del7697 (lacIPOZY)del74 galU galK hsdR  
rpsL srl recA56

Any other standard cloning recipient such as HB101 would be adequate.

Example 2

The front end of the synthetic HRP gene prepared in Example 1 was modified by the replacement of the HindIII-HpaI fragment with a synthetic linker carrying an NdeI site on the initiator ATG as follows:

M Q L T . . .  
 AAGCTTCATATGCAGTTAACC.....  
HindIII NdeI HpaI

Example 3

**Expression of the Synthetic Horseradish Peroxidase Gene in Escherichia coli**

The synthetic HRP gene of Example 2 was cloned into the expression vector pGC517 on a NdeI-BamHI fragment to give the plasmid pSD18. The host vector pGC517 was prepared from the known plasmid pAT153 (Twigg & Sherratt Nature 283, 216-218 (1980)), which is now a standard E. coli high expression vector, by the incorporation by standard methods of the known tac promoter sequence and a termination sequence. pAT153 is itself a derivative of pBR322. In pGC517 the HRP gene is expressed from the powerful and regulatable tac promoter. To ensure that expression remained repressed in uninduced cultures the plasmid was maintained in E. coli strain W3110 lacI<sup>q</sup>, which is widely available, in which the lac repressor protein is over-produced. Figure 5 depicts the structure of pSD18.

Strain W3110 lacI<sup>q</sup>-pSD18 was grown in M9 minimal medium containing 0.2% glucose and 0.2% casamino acids. At an O.D. of 0.2 - 0.3 the culture was induced by the

addition of IPTG to a final concentration of 5mM. The culture was grown for a further 3 hr with samples removed at 30 min intervals.

Microscopic examination of the induced culture revealed the presence of inclusion bodies, characteristic of the accumulation of large amounts of insoluble aggregated protein within the cell. In addition, cultures expressing HRP at high levels acquired a pink colouration, perhaps related to the overexpression of a haem protein. SDS/PAGE analysis subsequently revealed the presence of a large amount of a 33 kD protein, estimated at 10-20% of total cell protein in induced but not uninduced cultures. Western blot analysis confirmed that this protein was HRP.

Standard methods for inclusion body isolation could be applied to obtain a substantial purification of the denatured HRP as insoluble aggregates. This material was then dissolved in 6 M guanidine HCl prior to renaturation. For renaturation, the dissolved HRP was dialysed against 8 M urea, 50mM Tris HCl, 100mM NaCl for 24 hr.  $\text{Ca}^{2+}$  was then added (as  $\text{CaCl}_2$ ) to 1 mM and the sample incubated for 2 hr at room temperature. This procedure resulted in the recovery of about 0.125% of the expected HRP activity by the standard pyrogallol colorimetric assay and based on the protein concentration and estimated purity of the preparation (see Table 1).

Table 1 - Renaturation of HRP Expressed in E. coli

S a m p l e	Conditions	Rate of reaction (maximum) AU/min	Amount of recombinant HRP C assayed mcg	Activity AU/min mcg rec. HRP	Activity (% of max. activity of commercial HRP)
1	before 1st dialysis	0.01 AU/0.8 min	25 mcg	$5 \times 10^{-3}$ AU/min mcg	0.007%
2	after 1st dialysis	0.015 AU/1.1 min	5.77 mcg	0.0024 AU/min mcg	0.034%
3	sample 2 incubated with 1 mM Ca <sup>2+</sup> for 2h	0.01 AU/1.5 min	0.76 mcg	0.029 AU/min mcg	0.125%

Control samples prepared from similar cultures carrying the expression plasmid without the HRP gene gave backgrounds about 1000 fold less than this. The assay mixture contained freshly prepared pyrogallol and peroxide in the following concentrations: 11mM K phosphite, pH 6.0, 8mM H<sub>2</sub>O<sub>2</sub>, 0.55% w/v pyrogallol in H<sub>2</sub>O. The HRP was added and the increase in adsorption at 420nm was followed.

Thus the synthetic HRP gene is capable of high level expression in E. coli and is capable of directing the synthesis of active product.

#### Example 4

The synthetic HRP gene of Example 2 was modified as follows to allow for its efficient expression in mammalian cells:

(a) The 3' end of the gene was extended from the Pst 1 site to include the C-terminal extension reported by Fujiyama *et al.* Eur. J. Biochem. 173, 681-687 (1988).

(b) The 5' end of the gene was modified by the addition of a HindIII/HpaI linker which encoded a signal sequence based on an immunoglobulin signalpeptide.

The modified HRP gene is depicted in Figure 6, and will be referred to as HPRX.

#### Example 5

##### Expression of the Synthetic Horesradish Peroxidase Gene in Mammalian Cells

The HRPX gene of Example 4 was inserted into the mammlian cell expression vector pCPH11 to give pCP21, in which the HRP gene is expressed from the HCMV (Human Cytomegalovirus) early promoter, see Figure 7. The plasmid pCPH11 is based on pUC18, which is widely avaiialble and from which it can be prepared by standard methods, using the information in Figure 7.

The HRP expression plasmid pCP21 was transfected into COS cells using the standard technique of calcium phosphate precipitation (20mcg DNA transfected per  $10^6$



cells). HRP activity was assayed in cell culture medium, 48-72h post transfection using tetra-methyl benzidine substrate (TMB), a standard HRP reagent. No HRP activity was detectable in control constructs which did not contain a signal sequence and/or the 3' extension. In contrast, HRP activity was clearly detectable in cells transfected with pCP21 (up to 10x greater than in controls). The results are shown in Table 2.

Table 2 - HRP Expression in COS Cells

Vol.Extract. (mcl)	O.D.450			
	<u>Plasmid</u>			
	<u>pCP21</u>	<u>pCP22</u>	<u>pCP11</u>	<u>pCP12</u>
100	.004	.022	.008	.003
50	.033	.012	.010	.013
25	.107*	.010	.003	.016
10	.084*	.011	.007	.017
5	.064*	.015	.012	.011
1	.028	.008	.007	.007

KEY

pCP21	HRP with N and C terminal signals, correct orientation.
pCP22	HRP with N and C terminal signals, wrong orientation.
pCP11	HRP with no signal sequences, correct orientation.
pCP12	HRP with no signal sequences, wrong orientation.

All results are the mean of duplicate samples.

\* = significant level of activity.

#### HRP Assay

For assaying cell extracts, a substrate mix was prepared as follows:

TMB (3,3',5,5' tetramethyl benzidine (Sigma)) was dissolved to 10 mg/ml in DMSO and 100 mcl of this solution added to 100 ml of assay buffer (0.1M NaAc in citric acid, pH6.0) along with 100 mcl H<sub>2</sub>O<sub>2</sub>.

A cell extract was prepared by collecting the cells by centrifugation followed by freeze thawing or sonication. The medium, cell lysates and standards were aliquoted in 96 well microtitre plates as follows:

Sample	100	50	25	10	5	1	mcl
Assay	0	50	75	90	95	99	mcl
Buffer							

Blank samples were set up using 100 mcl of assay buffer alone. 100 mcl of TMB/H<sub>2</sub>O<sub>2</sub> mix was added to the samples of incubated at RT for 30 mins to 1 hour. The reaction was stopped by the addition of 50 mcl of 2.5M H<sub>2</sub>SO<sub>4</sub> and the colour change read at 450 nm on a plate reader.

Commercially available HRP was used as a standard diluted by a factor of 10<sup>-6</sup>.

CLAIMS

1. DNA coding for horseradish peroxidase and having restriction sites for the following enzymes:

HpaI, SacI, SspI, NheI, NsiI, ClaI, PvuI, SphI,  
BspMI, FspI, RsrII, SalI, BglII, BalI, PvuII,  
XhoI, BspMII, BstEII, SnaBI, PflMI, ApaLI, SpeI,  
ScaI, XbaI, StuI, BclI, BstXI, NcoI, KpnI and PstI

2. DNA including the following sequence:

CAG TTA ACC CCT ACA TTC TAC GAC AAT AGC TGT CCC  
AAC GTG TCC AAC ATC GTT CGC GAC ACA ATC GTC AAC  
GAG CTC AGA TCC GAT CCC AGG ATC GCT GCT TCA ATA  
TTA CGT CTG CAC TTC CAT GAC TGC TTC GTG AAT GGT  
TGC GAC GCT AGC ATA TTA CTG GAC AAC ACC ACC AGT  
TTC CGC ACT GAA AAG GAT GCA TTC GGG AAC GCT AAC  
AGC GCC AGG GGC TTT CCA GTG ATC GAT CGC ATG AAG  
GCT GCC GTT GAG TCA GCA TGC CCA CGA ACA GTC AGT  
TGT GCA GAC CTG CTG ACT ATA GCT GCG CAA CAG AGC  
GTG ACT CTT GCA GGC GGA CCG TCC TGG AGA GTG CCG  
CTC GGT CGA CGT GAC TCC CTA CAG GCA TTC CTA GAT  
CTG GCC AAC GCC AAC TTG CCT GCT CCA TTC TTC ACC  
CTG CCC CAG CTG AAG GAT AGC TTT AGA AAC GTG GGT  
CTG AAT CGC TCG AGT GAC CTT GTG GCT CTG TCC GGA  
GGA CAC ACA TTT GGA AAG AAC CAG TGT AGG TTC ATC  
ATG GAT AGG CTC TAC AAT TTC AGC AAC ACT GGG TTA  
CCT GAC CCC ACG CTG AAC ACT ACG TAT CTC CAG ACA  
CTG AGA GGC TTG TGC CCA CTG AAT GGC AAC CTC AGT  
GCA CTA GTG GAC TTT GAT CTG CGG ACC CCA ACC ATC  
TTC GAT AAC AAG TAC TAT GTG AAT CTA GAG GAG CAG  
AAA GGC CTG ATA CAG AGT GAT CAA GAA CTG TTT AGC

AGT CCA AAC GCC ACT GAC ACC ATC CCA CTG GTG AGA  
AGT TTT GCT AAC TCT ACT CAA ACC TTC TTT AAC GCC  
TTC GTG GAA GCC ATG GAC CGT ATG GGT AAC ATT ACC  
CCT CTG ACG GGT ACC CAA GGC CAG ATT CGT CTG AAC  
TGC AGA GTG GTC AAC AGC AAC TCT

3. A genetic construct comprising DNA as claimed in claim 1 or 2, or a fragment thereof.
4. A construct as claimed in any one of claims 3 to 8, which is a vector.
5. A process for the preparation of DNA as claimed in claim 1 or 2 or a genetic construct in accordance with claim 3, the process comprising coupling successive nucleotides and/or ligating appropriate oligomers.
6. DNA substantially as herein described with reference to Figure 2a.
7. A construct as claimed in claim 3 comprising all or a fragment of the sequence defined in claim 1 or 2 fused to any other sequence of DNA so as to result in a sequence capable of encoding a hybrid protein possessing peroxidase activity.
8. A process for the production of monodisperse horseradish peroxidase C comprising the expression of at least part of a genetic construct as claimed in claim 3.
9. A construct which is a genetic fusion between a gene encoding horseradish peroxidase and a gene

encoding streptavidin or avidin such that the encoded fusion protein possesses both biotin binding and peroxidase activity.

10. A genetic fusion between a gene encoding horseradish peroxidase and a gene encoding an immunoglobulin-derived antigen binding function such that the fusion protein possesses both antigen binding and horseradish peroxidase activity.

11. A genetic fusion as claimed in claim 10, wherein the antigen binding function is an immunoglobulin heavy chain or light chain or fragments thereof or an engineered monomeric antigenic recognition site.

12. A construct as claimed in claim 4, which is a vector comprising the gene for horseradish peroxidase C that enables the production of fusions between horseradish peroxidase and any other protein of interest.

13. A construct as claimed in claim 4, which is an expression vector that provides for the co-expression of the gene for horseradish peroxidase and another gene of interest either as a single fusion product, as a single polycistronic message or as two separate but linked transcriptional units.

14. A gene for horseradish peroxidase containing a mutation (either missense, nonsense, deletion, insertion, duplication or other re-arrangement) that destroys or impairs the activity of the encoded horseradish peroxidase protein.

15. A genetic construct including all or a fragment of a mutant horseradish peroxidase gene as claimed in claim 14.

16. The use of a defective horseradish peroxidase genetic construct as claimed in claim 15 where the genetic construct is employed in mammalian cells.

17. The use of a defective horseradish peroxidase genetic construct as claimed in claim 15 where the genetic construct is employed in transgenic animals.

18. Any novel feature or combination of features disclosed herein (whether or not otherwise claimed).

## FIGURE 1

AMINO ACID SEQUENCE OF HORSERADISH PEROXIDASE (INCLUDING  
INITIATOR METHIONINE).

(M) Q L T P T F Y D N S C P N V S N I V R  
10 20  
D T I V N E L R S D P R I A A S I L R L  
30 40  
H F H D C F V N G C D A S I L L D N T T  
50 60  
S F R T E K D A F G N A N S A R G F P V  
70 80  
I D R M K A A V E S A C P R T V S C A D  
90 100  
L L T I A A Q Q S V T L A G G P S W R V  
110 120  
P L G R R D S L Q A F L D L A N A N L P  
130 140  
A P F F T L P Q L K D S F R N V G L N R  
150 160  
S S D L V A L S G G H T F G K N Q C R F  
170 180  
I M D R L Y N F S N T G L P D P T L N T  
190 200  
T Y L Q T L R G L C P L N G N L S A L V  
210 220  
D F D L R T P T I F D N K Y Y V N L E E  
230 240  
Q K G L I Q S D Q E L F S S P N A T D T  
250 260  
I P L V R S F A N S T Q T F F N A F V E  
270 280  
A M D R M G N I T P L T G T Q G Q I R L  
290 300  
N C R V V N S S  
309

FIGURE 2a

2/15

## SEQUENCE OF SYNTHETIC HORSERADISH PEROXIDASE GENE

```

      M Q L T P T F Y D N S C P N V S N
AAGCTTAACCATGCACTTAACCCCTACATTCTAAGACAATAGCTGTCCCAACGTTGTCAC
HindIII      HpaI
TTGGAATTGGTACGTCAATTGGGGATGTAAGATGCTGTATATGACAGGGTTGCACAGGTT
      10      20      30      40      50      60

      I V R D T I V N E L R S D P R I A A S I
CATGGTTGCGACACAATGCTCAACGAGCTCAGATCCGATCCAGGATGCTGCTTCAAT
      NruI      SacI      SspI
GTAGCAAGCGCTGTGTAGCAGTTGCTGAGTCTAGGCTAGGGTCTAGGAGCAAGTTA
      70      80      90      100     110     120

      L R L H F H D C F V N G C D A S I L L D
ATTACGTCGCACTTCCATGACTGCTTGGTGAATGGTTGCGAAGCTAGCATATTACTGGA
      NheI
TAATGCAGACGTGAAGGTACTGACGAAGCACTTACCAACGCTGCGATGTATAATGACCT
      130     140     150     160     170     180

      N T T S F R T E K D A F G N A N S A R G
CAACACCACAGTTTCCGCACTGAAAAGGATGCATTGGGAAAGCTAACAGCGCCAGGGG
      NsiI
GTTGTGGTGGTCAAAGGCGTACCTTTTCTACGTAAGCCCTTGGGATTGTGCGGGTCCCC
      190     200     210     220     230     240

      F P V I D R M K A A V E S A C P R T V S
CTTTCCAGTACGATGCGATGAAGGCTGCGGTGAGTCAGCATGCCACGAACAGTCAG
      ClaI/PvuI      SphI
GAAAGGTACCTAGCTAGGTTACCTTCCGACGGCACTCAGTGTACGGGTGCTTGTTCAGTC
      250     260     270     280     290     300

      C A D L L T I A A Q Q S V T L A G G P S
TTGTGCAGACCTGCTGACTATAGCTGCGCAACAGAGCGTGAATCTTGCAGGCGGACGTC
      BspMI      FspI      RsrII
AACACGTCCTGGACGACTGATATGACGCGGTTGTCTGCACTGAGAACGTCGCGCTGGCAG
      310     320     330     340     350     360

      W R V P L G R R D S L Q A F L D L A N A
CTGGAGAGTGGCGCTCGGTGACGTGACTCCCTACAGGCATTCCCTAGATCTGGCCAACGC
      SalI      BglII/BalI
GACCTCTCAGCGGAGCGAGCTGCACTGAGGGATGTCCGTAAAGGATCTAGACGGTTGGG
      370     380     390     400     410     420

      N L P A P F F T L P Q L K D S F R N V G
CAACTTGCGTCTCCATTCTTCACCCCTGCCCCAGCTGAAGGATAGCTTTAGAAACGTTGG
      PvuII
GTTGAACCGACGAGGTAAGAAGTGGGACGGGGTGGACTTCCTATCGAAATCTTTGCACCC
      430     440     450     460     470     480

      L N R S S D L V A L S G G H T F G K N Q
TCTGAATGCTCGAGTGACCTTGTGGCTCTGTCCGGAGGACACATTGTGAAAGAACCA
      XhoI      BspMII
AGACTTAGCGAGCTCACTGGAACACCGAGACAGGCCTCTGTGTGTAAACCTTTCTTGGT
      490     500     510     520     530     540

```



3/15

C R F I M D R L Y N F S N T G L P D P T  
G T G T A G G T T C A T C A T G G A T A G G C T C T A C A A T T T C A G C A A C A C T G G G T T A C C T G A C C C C A C

BstEII

C A C A T C C A A G T A G T A C C T A T C C G A G A T G T T A A A G T O G T T G T G A C C C A A T G G A C T G G G G T G  
550 560 570 580 590 600

L N T T Y L Q T L R G L C P L N G N L S  
G C T G A A C A C T A C G T A T C T C C A G A C A C T G A G A G G C T T G T G C C C A C T G A A T G G C A A C C T C A G

SnaBI

PflMI

O G A C T T G T G A T G C A T A G A G G T C T G T G A C T C T C C G A A C A C G G G T G A C T T A C O G T T G G A G T C  
610 620 630 640 650 660

A L V D F D L R T P T I F D N K Y Y V N  
T G C A C T A G T G G A C T T T G A T C T G C G G A C C C A A C C A T C T T C G A T A A C A A G T A C T A T G T G A A

ApaLI/SpeI

ScaI

A O G T G A T C A C C T G A A A C T A G A C G C C T G G G G T T G G T A G A A G C T A T T G T T C A T G A T A C A C T T  
670 680 690 700 710 720

L E E Q K G L I Q S D Q E L F S S P N A  
T C T A G A G G A G C A G A A A G G C C T G A T A C A G A G T G A T C A A G A A C T G T T T A G C A G T C C A A A A G C

XbaI

StuI

BclI

A G A T C T C C T O G T C T T T C C G G A C T A T G T C T C A C T A G T T C T T G A C A A A T O G T C A G G T T T G C G  
730 740 750 760 770 780

T D T I P L V R S F A N S T Q T F F N A  
C A C T G A C A C C A T C C C A C T G G T G A G A A G T T T T G C T A A C T C T A C T C A A A C C T T C T T T A A C G C

BstXI

G T G A C T G T G G T A G G G T G A C C A C T C T T C A A A A C G A T T G A G A T G A G T T T G G A A G A A A T T G C G  
790 800 810 820 830 840

F V E A M D R M G N I T P L T G T Q G Q  
C T T O G T G G A A G C C A T G G A C C G T A T G G G T A A C A T T A C C C C T C T G A C G G G T A C C C A A G G C C A

NcoI

KpnI

G A A G C A C C T T O G G T A C C T G G C A T A C C A T T G T A A T G G G G A G A C T G C C C A T G G G T T C O G G T  
850 860 870 880 890 900

I R L N C R V V N S N S \* \*  
G A T T O G T C T G A A C T G C A G A T G G T C A A C A G C A A C T C T T A A T A A G G A T C C G A A T T C

PstI

BamHI EcoRI

C T A A G C A G A C T T G A C G T C T A C C A G T T G T O G T T G A G A A T T A T T C T A G G C T T A A G  
910 920 930 940 950

4/15

FIGURE 2b  
SUMMARY OF USEFUL RESTRICTION SITES.

ENZYME	SEQUENCE	POSITION
HindIII	AAGCTT	1
HpaI	GTTAAC	16
SacI	GAGCTC	86
SspI	AATATT	118
NheI	GCTAGC	164
NsiI	ATGCAT	210
ClaI	ATCGAT	251
PvuI	CGATCG	253
SphI	GCAATG	281
BspMI	ACCTGC	309
FspI	TGCGCA	325
RsrII	CGGACCG	352
SalI	GTCGAC	378
BglII	AGATCT	406
BalI	TGGCCA	411
PvuII	CAGCTG	452
XhoI	CTCGAG	490
BspMII	TCGGGA	512
BstEII	GGTTACC	585
SnaBI	TACGTA	610
PfIMI	CCACTGAATGG	641
ApaLI	GTGCAC	660
SpeI	ACTAGT	664
ScaI	AGTACT	708
XbaI	TCTAGA	721
StuI	AGGCCT	736
BclI	TGATCA	751
BstXI	CCATCCCACTGG	789
NcoI	CCATGG	852
KpnI	GGTACC	887
PstI	CTGCAG	913
BamHI	GGATCC	944
EcoRI	GAATTC	950

5/15

FIGURE 2c

SEQUENCE OF SYNTHETIC HORSE RADISH PEROXIDASE GENE 5' END HALF

M Q L T P T F Y D N S C P N V S N  
 AAGCTTAACCATGCGAGTTAACCCCTACATTCACGACAATAGCTGTCCCAACGTGTCCAA  
 HindIII HpaI  
 TTCGAATTGGTAAGTCAATTGGGGATGTAAGATGCTGTTATOGACAGGGTTGCACAGTT  
 10 20 30 40 50 60

I V R D T I V N E L R S D P R I A A S I  
 CATOGTTGCGACACAATOGTCAACGAGCTCAGATCOGATCCAGGATOGCTGCTTCAAT  
 NruI SacI SspI  
 GTAGCAAGOGCTGTGTTAGCAGTTGCTOGAGTCTAGGCTAGGGTCTAGOGAOGAAGTTA  
 70 80 90 100 110 120

L R L H F H D C F V N G C D A S I L L D  
 ATTAAGTCTGCACCTCCATGACTGCTTGTGTAATGGTTGOGAGCTAGCATATTACITGA  
 NheI  
 TAATGCAGAOGTGAAGGTACTGAOGAAGCACTTACCAACGCTGOGATOGTATAATGAOCT  
 130 140 150 160 170 180

N T T S F R T E K D A F G N A N S A R G  
 CAACACCACAGTTTCCGCACTGAAAAGGATGCATTGGGAACGCTAACAGOGCCAGGGG  
 NsiI  
 GTTGTGGTGGTCAAAGGOGTGACTTTTCTAGCTAAGCCCTTGGGATTGTGGGGTCCCC  
 190 200 210 220 230 240

F P V I D R M K A A V E S A C P R T V S  
 CTTTCCAGTGTATGATGCGATGAAGGCTGCGGTTGAGTCAGCATGCCAGGAACAGTCAG  
 ClaI/PvuI SphI  
 GAAAGGTCACTAGCTAGCGTACTTCCGAAGGCAACTCAGTGTAGCGGGTGTGTGTCAGTC  
 250 260 270 280 290 300

C A D L L T I A A Q Q S V T L A G G P S  
 TTGTGCAGAOCTGCTGACTATAGCTGCGCAACAGAGOGTGACTCTTGCAGGCGGACCGTC  
 BspMI FspI RsrII  
 AACAGCTCTGGAAGACTGATATGCAAGCGTTGTCTCGCACTGAGAACGTCCGCGCTGGCAG  
 310 320 330 340 350 360

W R V P L G R R D S L Q A F L D L A N A  
 CTGGAGAGTGGCGCTGGTGGAGCTGCTGCTACAGGCATTCTAGATCTGGCCAAOGC  
 Sali BglII/BalI  
 GAOCCTCTCAAGGOGAGCCAGCTGCACTGAGGGATGTCCGTAAGGATCTAGACOGGTTGGG  
 370 380 390 400 410 420

N L P A P F F T L P Q L K D S F R N V G  
 CAACTTGCCCTGCTCCATTCTTCACCCCTGCCCGAGCTGAAGGATAGCTTTAGAAACGTGGG  
 PvuII  
 GTTGAACGGAAGGTAAGAAGTGGGAAGGGTGGAGTCTTCCCTATOGAAATCTTTGCACCC  
 430 440 450 460 470 480

L N R S  
 TCTGAATGCTCGAGGAATTC  
 XhoI EcoRI  
 AGACTTAGCGAGCTCCCTAAG  
 490

FIGURE 2d

6/15

SEQUENCE OF SYNTHETIC HORSE RADISH PEROXIDASE GENE 3' END HALF.

S S D L V A L S G G H T F G K N Q  
 AAGCTTAACCTGAGTGAACCTTGTGGCTCTGTCCGGAGGACACACATTTGGAAAGAACCA  
 HindIII XhoI BspMII  
 TTGGAATTGAGCTCACTGGAACACCGAGACAGGCTCTGTGTGTAAACCTTTCCTGGT  
 490 500 510 520 530 540

C R F I M D R L Y N F S N T G L P D P T  
 GTGTAGGTTTCATCATGGATAGGCTCTACAATTTGAGCAACACTGGGTTACCTGACCCAC  
 BstEII  
 CACATCCAAGTAGTACCTATCCGAGATGTTAAAGTGGTTGTGACCCAATGGACTGGGGTG  
 550 560 570 580 590 600

L N T T Y L Q T L R G L C P L N G N L S  
 GCTGAACACTAGCTATCTCCAGACACTGAGAGGCTGTGCCCCTGAATGGCAACCTCAG  
 SnaBI PflMI  
 CGACTTGTGATGCATAGAGGCTCTGTGACTCTCCGAACACGGGTGACTTACCGTTGGAGTC  
 610 620 630 640 650 660

A L V D F D L R T P T I F D N K Y Y V N  
 TGCCTAGTGGACTTTGATCTGCGGACCCCAACCATCTTGATAACAAGTACTATGTGAA  
 ApaLI/SpeI ScaI  
 ACGTGATCACTGAAACTAGACGCTGGGGTTGGTGAAGCTATTGTTTCATGATACACTT  
 670 680 690 700 710 720

L E E Q K G L I Q S D Q E L F S S P N A  
 TCTAGAGGAGCAGAAAGGCTGATACAGAGTGATCAAGAAGCTGTTAGCAGTCCAAACGC  
 XbaI StuI BclI  
 AGATCTCTCTGCTTTCCGACTATGTCTCACTAGTTCTTGACAAATGCTCAGGTTTGGC  
 730 740 750 760 770 780

T D T I P L V R S F A N S T Q T F F N A  
 CACTGACACCATCCACTGGTGAGAAGTTTGTCTAACTCTACTCAAACCTTCTTTAACGC  
 BstXI  
 GTGACTGTGGTAGGGTGACCACTCTTCAAAACGATTGAGATGAGTTTGAAGAAATTGGC  
 790 800 810 820 830 840

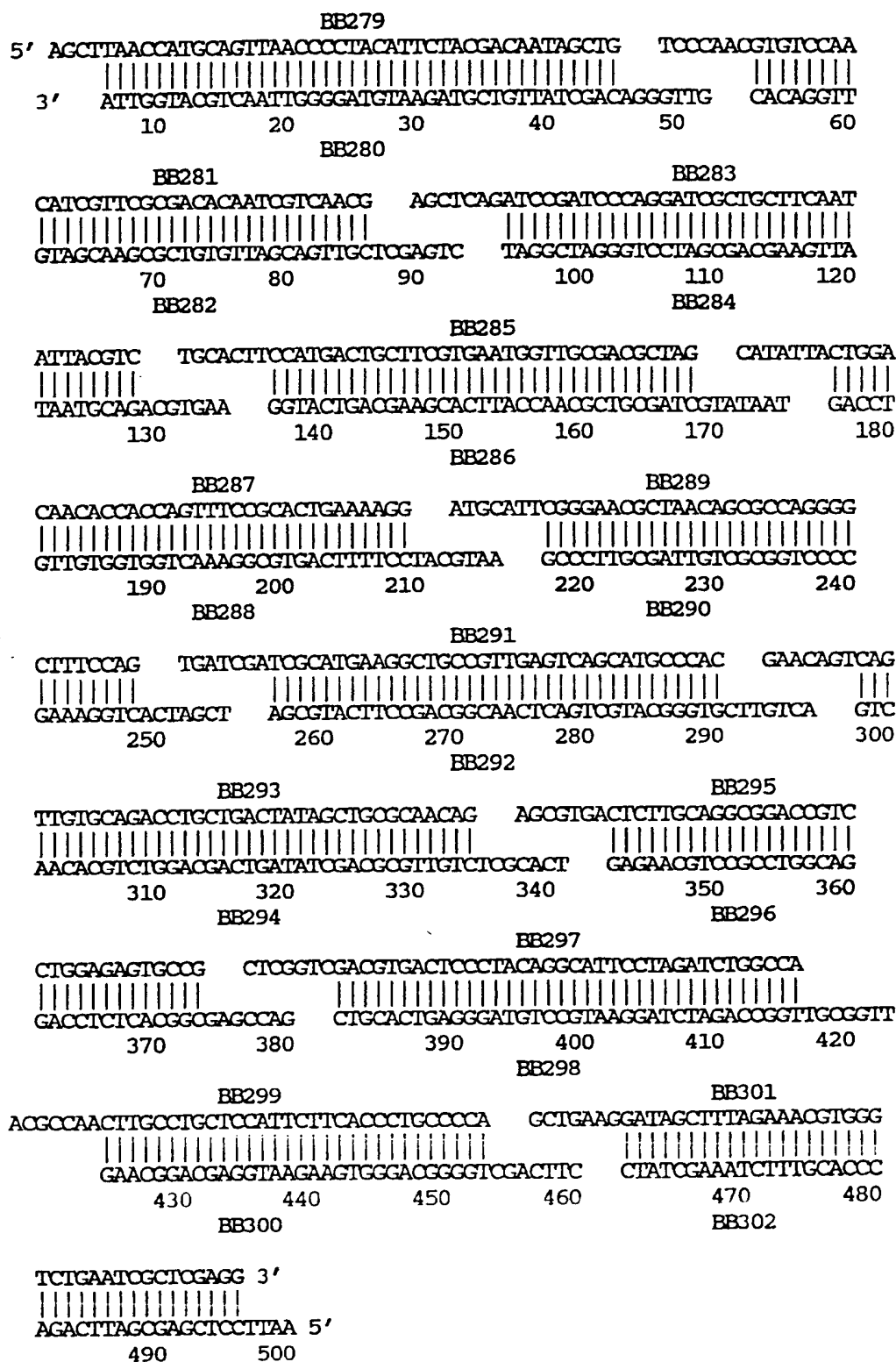
F V E A M D R M G N I T P L T G T Q G Q  
 CTTGTGGAAGCATGGACGATGGGTAAACATTACCCCTCTGACGGGTACCCAAGGCCA  
 NcoI KpnI  
 GAAGCACCTTGGTAACCTGGCATAACCAATTGTAATGGGGAGACTGCCCATGGGTTCCGGT  
 850 860 870 880 890 900

I R L N C R V V N S N S \* \*  
 GATTGCTCTGAACCTGCAGAGTGGTCAACAGCAACTCTTAATAAGGATCCGAATTC  
 PstI BamHI EcoRI  
 CTAAGCAGACTTGAAGTCTCAACAGTTGTGGTTGAGAATTATTCCTAGGCTTAAG  
 910 920 930 940 950

FIGURE 3a

7/15

DESIGN OF OLIGOMERS FOR SYNTHETIC HORSE RADISH PEROXIDASE GENE 5' HALF.



8/15

FIGURE 3b

DESIGN OF OLIGOMERS FOR SYNTHETIC HORSERADISH PEROXIDASE GENE 3' HALF.

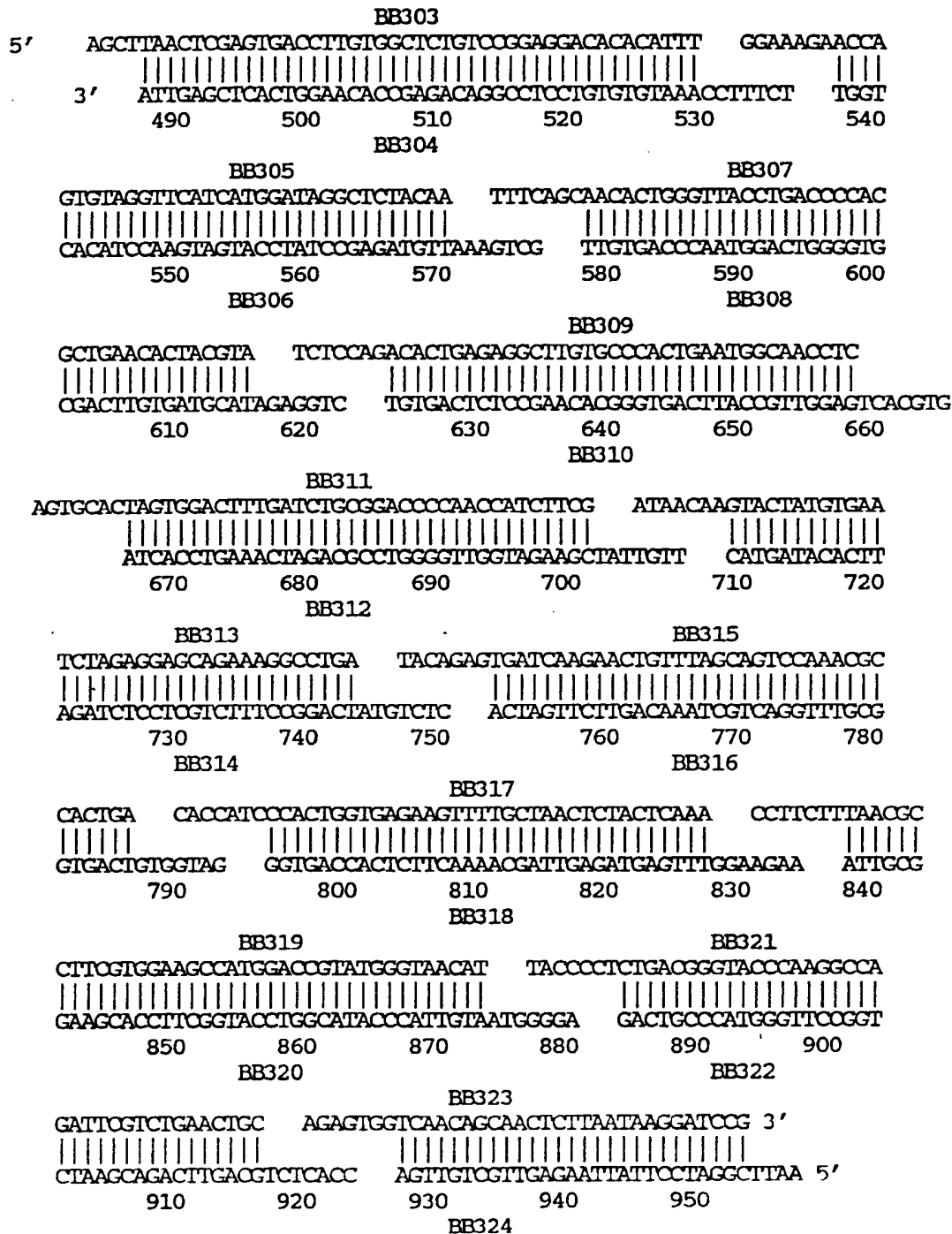
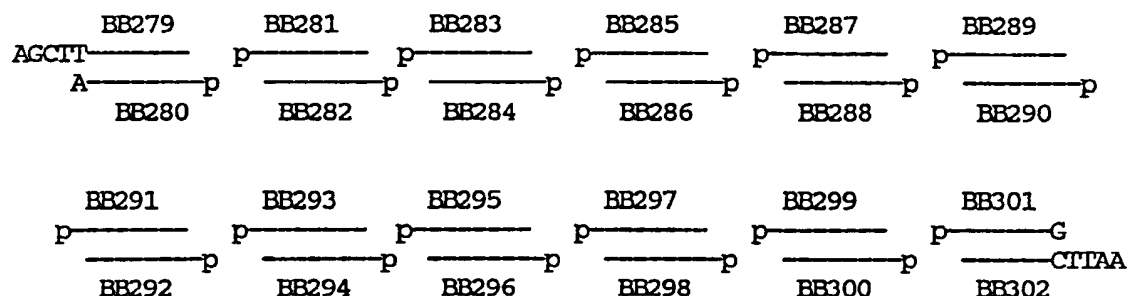


FIGURE 4a

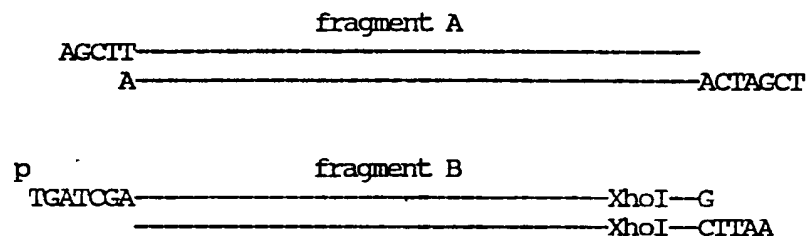
9/15

## SUMMARY OF ASSEMBLY PROCEDURE, 5' HALF.

- a) kinased oligomers annealed in pairs and mixed in two groups (A & B).



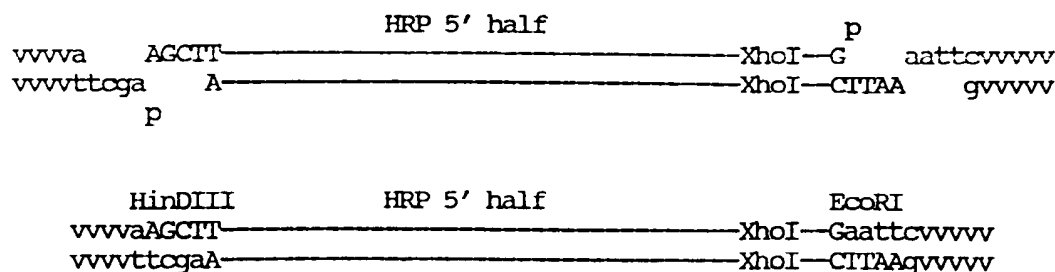
- b) oligomers ligated together in two groups. (BB279 and BB302 not kinased to avoid multimerisation).



- c) The ligations were checked for the presence of fragment A & B on 2% agarose gels then the ligation reactions were mixed and the reaction allowed to continue to give the final product.



- d) The HRP gene fragment was isolated on a 2% LGT agarose gel and cloned into EcoRI/HindIII cut pUC18.



v = vector sequence

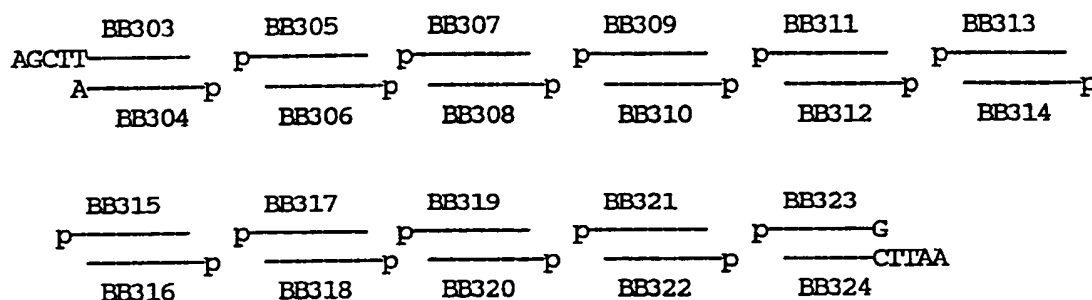
p = 5' phosphates

FIGURE 4b

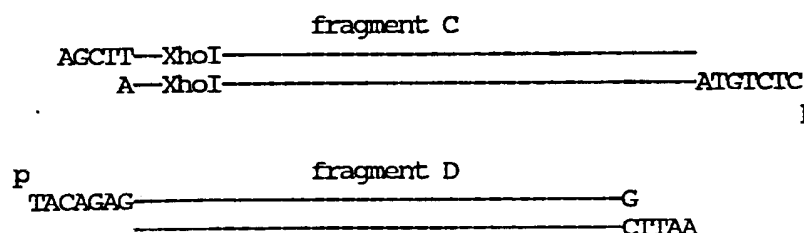
10/15

## SUMMARY OF ASSEMBLY PROCEDURE, 3' HALF.

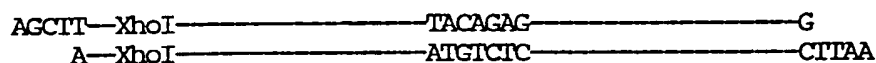
- a) kinased oligomers annealed in pairs and mixed in two groups (A & B).



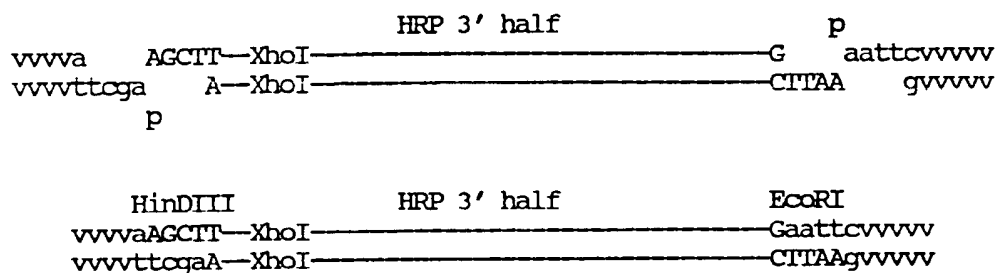
- b) oligomers ligated together in two groups. (BB303 and BB324 not kinased to avoid multimerisation).



- c) The ligations were checked for the presence of fragment C & D on 2% agarose gels then the ligation reactions were mixed and the reaction allowed to continue to give the final product.



- d) The HRP gene fragment was isolated on a 2% IGT agarose gel and cloned into EcoRI/HindIII cut pUC18



v = vector sequence

p = 5' phosphates



11/15

## FIGURE 4c:

## SUMMARY OF FINAL ASSEMBLY PROCEDURE.

- a) 5' and 3' clones of HRP cloned in pUC18 were digested with XhoI and EcoRI. Relevant fragments from each digest were isolated from a 0.8% LGT agarose gel.

HindIII	HRP 5' HALF	XhoI	EcoRI
vvvvvAAGCTT	—————	CTGAG	GAATTCvvvvv
vvvvvTTCGAA	—————	GAGCTC	CTTAAGvvvvv

HindIII	XhoI	HRP 3' HALF	EcoRI
vvvvvAAGCTT	CTGAG	—————	GAATTCvvvvv
vvvvvTTCGAA	GAGCTC	—————	CTTAAGvvvvv

- b) XhoI/EcoRI fragment carrying 3' half of HRP ligated into XhoI/EcoRI cut HRP 5' half clone.

		HRP 3' HALF	
	TCGAG	—————	G
	C	—————	CTTAA
HRP 5' HALF			
vvvvvAAGCTT	—————	C	AATTCvvvvv
vvvvvTTCGAA	—————	GAGCT	Gvvvvv

- c) Completed gene cloned in pUC18.

HRP 5' HALF		HRP 3' HALF	
vvvvvAAGCTT	—————	CTGAG	GAATTCvvvvv
vvvvvTTCGAA	—————	GAGCTC	CTTAAGvvvvv

v = vector sequence

12/15

Figure 5. The HRP Expression Plasmid pSD18.

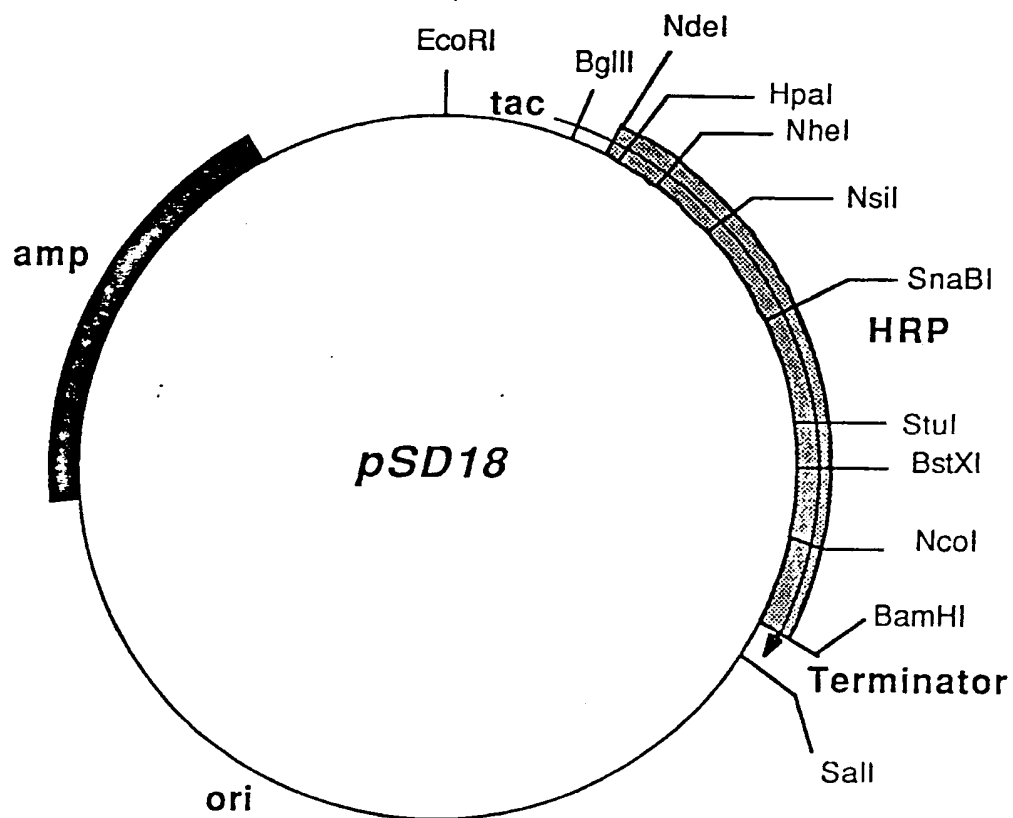


Figure 6. Synthetic HRP Gene Modified for Expression in Mammalian Cells.

```

      M k c s w v i f f l m a v v t g v
AAGCTTCCACCATGAAGTGCTCCTGGGTGATCTTCTTCCTGATGGCCGTGGTGACCGGCG
  10      20      30      40      50      60

      n s<>Q L T P T F Y D N S C P N V S N I V
TGAACTCCAGTTAACCCCTACATTCTACGACAATAGCTGTCCCAACGTGTCCAACATCG
  70HpaI  80      90      100      110      120

      R D T I V N E L R S D P R I A A S I L R
TTCGCGACACAATCGTCAACGAGCTCAGATCCGATCCCAGGATCGCTGCTTCAATATTAC
  130      140      150      160      170      180

      L H F H D C F V N G C D A S I L L D N T
GTCTGCACTTCCATGACTGCTTCGTGAATGGTTGCGACGCTAGCATATTACTGGACAACA
  190      200      210      220      230      240

      T S F R T E K D A F G N A N S A R G F P
CCACCAGTTTCCGCACTGAAAAGGATGCATTCGGAACGCTAACAGCGCCAGGGGCTTTC
  250      260      270      280      290      300

      V I D R M K A A V E S A C P R T V S C A
CAGTGATCGATCGCATGAAGGCTGCCGTTGAGTCAGCATGCCCACGAACAGTCAGTTGTG
  310      320      330      340      350      360

      D L L T I A A Q Q S V T L A G G P S W R
CAGACCTGCTGACTATAGCTGCGCAACAGAGCGTGACTCTTGCAAGCGGACCGTCCTGGA
  370      380      390      400      410      420

      V P L G R R D S L Q A F L D L A N A N L
GAGTGCCGCTCGGTGCGACGTGACTCCCTACAGGCATTCCTAGATCTGGCCAACGCCAACT
  430      440      450      460      470      480

      P A P F F T L P Q L K D S F R N V G L N
TGCCTGCTCCATTCTTCACCCTGCCCCAGCTGAAGGATAGCTTTAGAAACGTGGGTCTGA
  490      500      510      520      530      540

      R S S D L V A L S G G H T F G K N Q C R
ATCGCTCGAGTGACCTTGTGGCTCTGTCCGGAGGACACACATTGGAAGAACCAGTGTA
  550      560      570      580      590      600

      F I M D R L Y N F S N T G L P D P T L N
GGTTCATCATGGATAGGCTCTACAATTCAGCAACACTGGGTTACCTGACCCACGCTGA
  610      620      630      640      650      660

      T T Y L Q T L R G L C P L N G N L S A L
ACACTACGTATCTCCAGACACTGAGAGGCTTGTGCCCCTGAATGGCAACCTCAGTGCAC
  670      680      690      700      710      720

```

14/15

V D F D L R T P T I F D N K Y Y V N L E  
TAGTGGACTTTGATCTGCGGACCCCAACCATCTTCGATAACAAGTACTATGTGAATCTAG  
730 740 750 760 770 780

E Q K G L I Q S D Q E L F S S P N A T D  
AGGAGCAGAAAGGCCTGATACAGAGTGATCAAGAAGTGTTCGAGTCCAAACGCCACTG  
790 800 810 820 830 840

T I P L V R S F A N S T Q T F F N A F V  
ACACCATCCCCTGGTGAGAAGTTTTGCTAACTCTACTCAAACCTTCTTTAACGCCTTCG  
850 860 870 880 890 900

E A M D R M G N I T P L T G T Q G Q I R  
TGGAAGCCATGGACCGTATGGGTAACATTACCCCTCTGACGGGTACCCAAGGCCAGATTC  
910 920 930 940 950 960

L N C R V V N S N S l l h d m v e v v d  
GTCTGAACTGCAGAGTGGTCAACAGCAACTCTCTACTCCATGATATGGTGGAGGTCGTTG  
PstI 980 990 1000 1010 1020

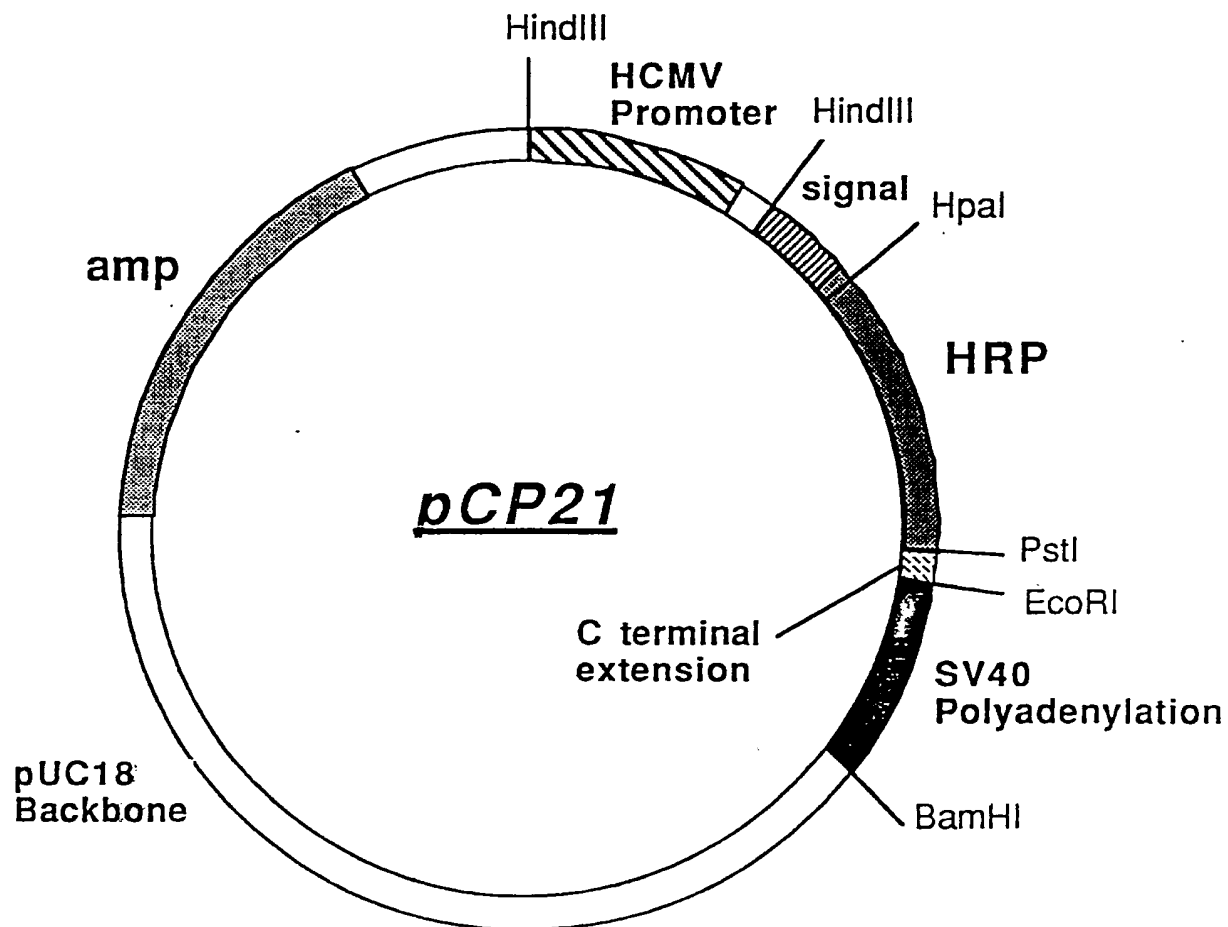
f v s s m \* \*  
ACTTTGTTAGCTCTATGTAATAAGGATCCGAATTC  
1030 1040 EcoRI

KEY

Underlined sequences indicate linkers used to adapt synthetic gene.

Lower case residues indicate N and C terminal pre & pro sequences.

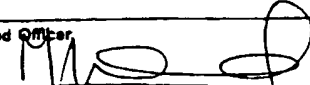
Figure 7.. The HRP Expression Plasmid pCP21.



# INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 88/00833

<b>I. CLASSIFICATION OF SUBJECT MATTER</b> (if several classification symbols apply, indicate all) <sup>6</sup>		
According to International Patent Classification (IPC) or to both National Classification and IPC		
IPC <sup>4</sup> :      C 12 N 15/00; C 12 N 9/08		
<b>II. FIELDS SEARCHED</b>		
Minimum Documentation Searched <sup>7</sup>		
Classification System	Classification Symbols	
IPC <sup>4</sup>	C 12 N	
Documentation Searched other than Minimum Documentation to the extent that such Documents are included in the Fields Searched <sup>8</sup>		
<b>III. DOCUMENTS CONSIDERED TO BE RELEVANT <sup>9</sup></b>		
Category <sup>9</sup>	Citation of Document, <sup>11</sup> with indication, where appropriate, of the relevant passages <sup>12</sup>	Relevant to Claim No. <sup>13</sup>
Y	European Journal of Biochemistry, volume 96, 1979, Springer-Verlag, (Berlin, DE), K.G. Welinder: "Amino acid sequence studies of horseradish peroxidase. Amino and carboxyl termini, cyanogen bromide and tryptic fragments, the complete sequence, and some structural characteristics of horseradish peroxidase C", pages 483-502 see the whole document cited in the application --	1-8
Y	EP, A, 0171024 (HOECHST AG) 12 February 1986 see page 4, lines 10-15 --	1-8
Y	EP, A, 0068375 (G.D. SEARLE & CO.) 5 January 1983 see page 4, line 13 - page 6a, line 4 -----	1-8
<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p><sup>10</sup> Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="width: 45%;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"A" document member of the same patent family</p> </div> </div>		
<b>IV. CERTIFICATION</b>		
Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report	
16th January 1989	- 2. 02. 89	
International Searching Authority	Signature of Authorized Officer	
EUROPEAN PATENT OFFICE	M. VAN MOL 	

# ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO.

GB 8800833  
SA 24660

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.  
The members are as contained in the European Patent Office EDP file on 23/01/89  
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A- 0171024	12-02-86	DE-A- 3429430	20-02-86
		AU-A- 4597785	13-02-86
		JP-A- 61047198	07-03-86
EP-A- 0068375	05-01-83	JP-A- 58010600	21-01-83
		AU-A- 8507282	06-01-83

EPO FORM P0479

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82